Target journal – currently formatted for Wildlife Society Bulletin

* Rangelands/Wildlife
  + **Wildlife Society Bulletin\* (**Seems like the best fit as it is methods heavy, and geared toward methods development for future monitoring)
  + Journal of Wildlife Management
  + Rangeland Ecology and Management
* Environmental monitoring
  + [Environmental Monitoring and Assessment](https://www.springer.com/journal/10661/): Traditionally focused on pollutants, but has expanded to included remote sensing applications, especially in agriculture.
  + Ecological Indicators
* UAS/Remote sensing
  + Journal of Applied Remote Sensing
  + Remote Sensing
  + Remote Sensing of Environment

12 Dec 2022

RH: Kearney et al. - Prairie dog mapping with deep learning

**Toward broad-scale mapping and characterization of prairie dog colonies from airborne imagery using deep learning**

Sean P. Kearney, USDA – ARS, 1701 Centre Ave. Fort Collins, CO 81301 USA

Lauren M. Porensky, USDA – ARS, 1701 Centre Ave. Fort Collins, CO 81301 USA

David J. Augustine, USDA – ARS, 1701 Centre Ave. Fort Collins, CO 81301 USA

David W. Pellatz, Thunder Basin Grasslands Prairie Ecosystem Association, 671 Steinle Road, Douglas, WY, 82633, USA

**Correspondence:** Sean P. Kearney, USDA – ARS, 1701 Centre Ave. Fort Collins, CO 81301 USA. Email: [sean.kearney@usda.gov](mailto:sean.kearney@usda.gov)

**ABSTRACT** Monitoring wildlife is fundamental to managing the health of rangelands but challenging to execute due the extensive and dynamic nature of these ecosystems. The black-tailed prairie dog (*Cynomys ludovicianus*) is considered both a keystone species of conservation concern and a pest and is emblematic of a rangeland species that is of high priority for monitoring but presents logistical challenges to do so cost-effectively using ground-based methods. In this study, we conducted a robust evaluation of the potential to use deep learning to detect prairie dog burrows from remotely sensed imagery acquired from unmanned aerial systems (UAS). We processed UAS imagery to create RGB, topographic position index (TPI) and normalized difference vegetation index (NDVI) products at varying spatial resolutions (2 – 30 cm). We then evaluated the minimum set of inputs and image resolution required to train a deep convolutional neural network (CNN) for burrow detection and then scale this up to identify entire colonies. We validated results at the scale of individual burrows, sub-colony burrow density and range-wide colony area using ground and digitized observations. We found that models performed well up to a spatial resolution of about 15 cm. The 2 cm imagery proved computationally impractical, but performance did not decline between 2 and 5 cm imagery. The top models always included TPI and the combination of RGB + TPI tended to perform best across spatial resolutions. Adding NDVI did not improve most models. At 5 cm resolution, the top models achieved high precision and recall for detecting individual burrows (F-score 0.84 - 0.87) and burrow density was strongly correlated with validation data (r = 0.94 - 0.95). In pastures with active colonies, overlap between predicted and ground delineated colonies was high (60 – 94%) and area estimates were within 25%. The CNN-based approach did not distinguish between currently active colonies and a colony that had recently become inactive due to a sylvatic plague (*Yersinia pestis*) epizootic during the study year. However, further analysis showed that CNN-derived burrow density was related to colony age and satellite-derived vegetation conditions in active colonies, and that the plagued colony deviated from this trend. We conclude that a deep learning algorithm can accurately detect prairie dog burrows from RGB and TPI imagery acquired at 5 – 10 cm, and possibly up to 15 cm, resolution. Our study shows that accurate and consistent scaling from individual burrows to entire colonies is likely achievable but warrants further research. Combining periodic mapping of burrow density using CNN and UAS imagery with near-real-time mapping of vegetation conditions with satellite imagery may help identify recent colony abandonment, despite ongoing presence of burrows.

**KEYWORDS** burrow detection, convolutional neural networks, *Cynomys ludovicianus,* deep learning, prairie dog colony, rangeland ecology, remote sensing

**INTRODUCTION**

Monitoring the presence and abundance of grassland wildlife is a critical and challenging component of managing rangeland health. The semi-arid climate typical of many rangelands results in highly dynamic and heterogeneous conditions across space and time, making frequent monitoring desirable. However, rangeland landscapes are often extensive and difficult to access, which makes frequent pedestrian ground measurements challenging, expensive and prone to bias or error. Rangeland managers are looking for new techniques, such as utilizing remotely sensed imagery, that can provide frequent, accurate and extensive monitoring coverage to complement and enhance ground-based efforts.

The black-tailed prairie dog (*Cynomys ludovicianus*) is emblematic of a rangeland species that is of high priority for monitoring but presents logistical challenges to do so cost-effectively using ground-based methods. Black-tailed prairie dogs are a keystone species in grasslands. They are ecosystem engineers – their burrows provide refugia for small mammals, burrowing owls (*Athene cunicularia*) and reptiles and amphibians (Kretzer & Cully 2001; Sidleet al. 2001; Lantz & Conway 2009) – and they are an essential prey source for several threatened and endangered species, including the black-footed ferret (*Mustela nigripes*) (Roelle et al.2006). Despite their importance for conservation, black-tailed prairie dogs are often viewed as pests due to their competition with livestock for forage (Derner et al. 2006; Detling 2006) and they are frequently exterminated by humans using poison and other methods. In addition to anthropogenic lethal control, black-tailed prairie dogs populations face threats from widespread habitat conversion (CITATION) and frequent epizootics of the non-native sylvatic plague (*Yersinia pestis*) introduced from Europe at the turn of the 20th century (CITATION?). Thus, the species, once widely distributed throughout the North American Great Plains, has seen a decline in population and distribution of 90-98% since European colonization.

In response to concerns about declining populations of black-tailed prairie dogs and associated species (e.g., black-footed ferrets), many jurisdictions regularly monitor prairie dog colonies to estimate their size, particularly on public lands. Other groups may monitor colonies to plan lethal control operations and minimize competition with livestock, especially on private lands. Accurate maps are highly desirable to facilitate communication between those seeking to keep prairie dogs on the landscape and those seeking to reduce their presence. Minimum colony area thresholds may be set to trigger plague control, shooting closures or increased monitoring (e.g., Colorado Parks and Wildlife, 2020), while maximum area thresholds may also exist to trigger lethal population control when competition with livestock is of concern (CITATION?). Current colony area estimates are often obtained by traversing colony perimeters with a handheld GPS unit or conducting aerial surveys (e.g., Colorado Parks and Wildlife, 2020; TBGPEA citation?, McDonald et al., 2015).

In addition to monitoring colony size, rangeland managers would ideally have information related to activity and population density at the colony scale and within colonies. Colonies with higher population densities may experience greater plague transmission (Cully and Williams, 2001), and within-colony data on population density could be used to improve the efficacy of both plague control measures, such as dusting for fleas. Information on population density could also inform projections of colony expansion (CITATION), which could help to target if, when and where to target population control measures. Such within-colony information is rarely available from existing survey methods (CITATION).

Remotely sensed images acquired from satellites, manned aircraft or unmanned aerial systems (UAS, i.e., drones) have been widely applied in environmental monitoring and show promise for monitoring prairie dog colonies using manual digitization (Sidle et al., 2002; McDonald et al., 2015) and automated techniques (Delparte et al., 2019, OTHERS?). The spatial resolution of imagery likely needs to be fine enough to discern the individual burrow entrances (hereafter called burrows) constructed by prairie dogs, typified by a hole, surrounded by a mounded patch of bare soil about 1 – 3 m in diameter. At coarse spatial scales, colonies are characterized by increased bare ground associated with burrow mounds and vegetation clipping by individual prairie dogs. However, bare ground can be widespread in rangelands due to a many other factors, especially at certain times of year. Even at relatively fine spatial scales (e.g., < 5 m), other ground features typified by a circular bare patch, such as anthills, are prevalent in some rangeland landscapes and may look like prairie dog burrows based solely on their spectral reflectance, despite having a very different topographic profile (i.e., mounded hills rather than burrows).

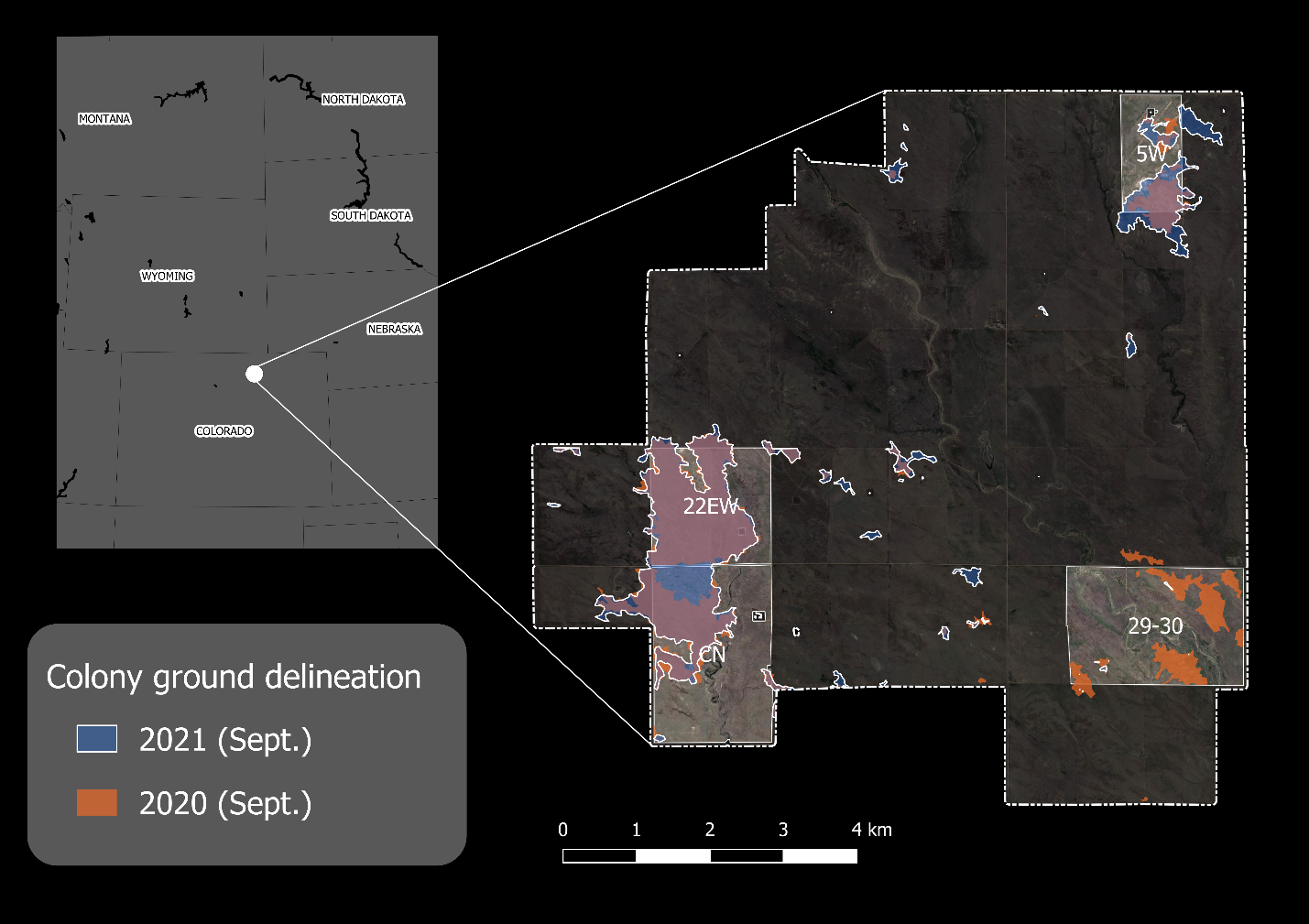
In this case study, we addressed several research objectives related to supporting prairie dog monitoring programs using remote sensing, with an eye toward extensive mapping of burrows and colonies in an accurate and repeatable manner. Specifically, we sought to (1) evaluate whether very high-resolution UAS imagery could be used to accurately detect individual black tailed prairie dog burrows, (2) evaluate whether a map of individual burrows could be scaled up to accurately delineate entire colonies and (3) identify the image inputs and minimum spatial resolution needed to achieve acceptable accuracies, since these will have a strong influence on cost and scalability.

We applied state of the art image acquisition and processing technologies and robust validation in this proof-of-concept study. For image acquisition, we used a lightweight fixed-wing UAS equipped with high resolution sensors to acquire imagery and a dense photogrammetric point cloud across a relatively large study area (1,120 ha). We employed deep learning image segmentation techniques to detect individual burrows. Deep neural networks have an uncanny ability to detect features within an image based on spatial patterning (CITATION), but to our knowledge have not yet been investigated to detect individual prairie dog burrows. We ran training and validation using different input combinations and artificially coarsened imagery to determine how each affected accuracy, and validated colony-scale predictions in two seasons to evaluate the transferability of models. We also evaluated how estimates of burrow density related to heterogeneity of colony age (using multi-year ground data of colony boundaries) and vegetation conditions (using satellite-derived maps of vegetation cover and biomass) within prairie dog colonies.

**STUDY AREA**

This study was carried out at the Central Plains Experimental Range (CPER) in NE Colorado, a US Department of Agriculture Long Term Agroecosystem Research (LTAR) site. [brief description of CPER. Mention prairie dog study?].

This study focused on four pastures encompassing 1,120 ha. Lethal prairie dog control is not used in these pastures, and they contained active colonies in 2020 and 2021 (Figure 1). Pasture 5W is characterized by [Describe veg and colony history...when was last plague]. Ground mapping indicated the colony was expanding from 2020 to 2021. Pastures 22EW and CN are separated by a gravel road, but share similar characteristics, with [describe vegetation and general colony history, last plague]. The colony boundaries of the colony spanning 22EW and CN were essentially unchanged between 2020 and 2021, however ground mapping indicated infilling on the north side of pasture CN. Pasture 29-30 is characterized by [describe veg]. A plague event occurred during the 2021 growing season. Ground mapping in 2020 and 2021 showed contraction of colony activity in the NE of the pasture in April 2021, and by September 2021 less than 1 ha was delineated as active.



**Figure 1.** The four study area pastures and ground delineated prairie dog colonies within the Central Plains Experimental Range (CPER) in NE Colorado. Dashed line shows the extent of CPER, and the four highlighted pastures were flown with a UAS in July and September of 2021. Orange polygons show colony delineation in September 2020 and blue polygons with white outlines show delineation in September 2021. Note that colonies were also delineated in April and July of 2021 (not shown).

**METHODS**

**Overview**

We iteratively tested the ability of a neural network to detect prairie dog burrows from UAS imagery by using different combinations of UAS image inputs (RGB, NDVI, TPI) at varying spatial resolutions (pixel sizes: 2 – 30 cm) for training the neural network. We conducted training at the pixel scale using semantic image segmentation and performed independent test validation at multiple scales. We first validated at the scale of individual burrows. Then, we converted predicted burrows to burrow density, which was independently validated at the scale of 30 x 30 m tiles. Finally, burrow density was used to classify entire prairie dog colonies, which were compared to ground-delineated colony perimeters to assess final area estimates and overlap of image-predicted vs ground-measured colonies. Finally, we analyzed heterogeneity within colonies by comparing variability of predicted burrow density to the variability of observed colony age and vegetation.

**Ground data**

Colony data

The delineation procedure followed methods employed at CPER annually beginning in 2000. Each year in September, all active prairie dog colonies at CPER were delineated using a handheld GPS unit. A colony was classified as active based on…

For the purpose of this study, we utilized existing ground data available from 2016 – 2020, and collected new data in 2021. During the 2021 study year, we surveyed colony boundaries in April and July, in addition to September, resulting in three separate colony boundary maps. The additional surveys were conducted to better understand change in colony activity throughout the season and enable validating our models in multiple seasons across varying vegetation conditions.

Plot data

During the summer of 2021, we set up 11 plots within they study area where we geolocated individual active prairie dog burrows and burrow-like feature (e.g., anthills, badger burrows, old and inactive prairie dog burrows) using a high precision GPS unit capable of sub-meter accuracy. For all active burrows, we also ranked (0 – 3) the size, height, activity and vegetation associated with each geolocated burrow. This ground data served three objectives. First, they provided reference points to train personnel for manual digitization of burrows from the imagery, which served as the training data for the deep neural network (see below). Second, a holdout of these ground data provided a validation dataset for assessing burrow-scale accuracy of both the manual digitization and the neural network output. Third, the holdout ground data allowed us to assess the rate of false positives associated with burrow-like features (e.g., anthills) and the characteristics (e.g., burrow size, activity level) of false negatives.

**UAS data**

Flight parameters

Image acquisitions were conducted in late July and early September 2021. An April acquisition was not feasible due to funding and logistical limitations. The two acquisitions were conducted using a Trinity F90+ fixed-wing UAS with vertical take-off and landing capability. During each acquisition, the entire area was covered twice, once with a Sony RX1 RII 42 MP camera (3-band RGB) and once with a Micasense RedEdge MX and MX Blue dual camera system (10-band multispectral). Flight altitude was limited to a maximum of 400 ft (122 m), resulting in a spatial resolution of 1.5-1.9cm for the RGB product and 8 – 10 cm for the multispectral product. The RGB flights had a 75% endlap and 70% sidelap, which produced a photgrammetric point cloud with about 220 points m-2, which was converted to a 6 cm digital surface model (DSM).

Image pre-processing

*[Could include a Paragraph from GTAC about geo-rectification, mosaicking, dense point cloud generation and DSM creation..but maybe overkill]*

From the DSM, we calculated a topographic position index (TPI) for two reasons: (1) to standardize these data for input into the deep neural network and (2) to highlight the burrow features, typified by a 10 – 30 cm hole surrounded by a mound up to 2.5 m in diameter (Cincotta 1989). We created the TPI by applying a moving window function using the Xarray-spatial package in Python. Specifically, for each pixel, we took the pixel’s elevation and from it subtracted the mean elevation of pixels in an annulus (i.e., doughnut) around the pixel. We set the annulus to have a minimum radius of 25 cm and a maximum radius of 75 cm, with the intention of capturing the mound surrounding the burrow. Therefore, negative TPI values indicate the surrounding elevation is higher than that of the pixel (i.e., the pixel is a local low point), positive values indicate the surrounding elevation is lower than that of the pixel (i.e., the pixel is a local high point) and values near zero indicate the pixel in a locally flat area.

We calculated the normalized difference vegetation index (NDVI) from the multispectral imagery as

where *NIR* is the near-infrared band, and *Red* is the red band of the visible spectra. The NDVI ranges from -1.0 to 1.0, with high values capturing green vegetation and low values indicative of bare soil. Negative values are rare, so we reset all negative values to zero, resulting a possible range of 0.0 – 1.0. We expected the NDVI to help highlight bare soil associated with active burrows and to help separate out old, inactive burrows that may have experienced recent revegetation.

**Burrow detection**

Data preparation for training and validation

To create a robust training and validation dataset, we manually digitized individual burrows by drawing polygons around burrows visible in the September imagery. We worked with the September imagery since annual colony We first created 30 x 30 m image tiles covering all the ground plots (n=45 tiles) with the geolocated burrows as well as a random sample of tiles generated at a rate of one per 10 ha across the study area (n=110). We then digitized all burrows visible in half the tiles covering the ground data and for all the randomly generated tiles, resulting in a total of 133 tiles. To train ourselves on how burrows and non-burrow features (e.g., anthills) appeared in the RGB, NDVI and TPI images, we inspected 50% of the tiles that covered the ground plots, with the ability to also visualize the ground data overlayed on the image layers. Of the randomly assigned tiles, we kept 80% (n=88 tiles) for training and set aside the other 20%, along with the digitized tiles covering the ground data, for testing (i.e., independent validation; n=45 tiles).

For all tiles, we extracted smaller image windows to serve as the final training and validation images for the neural network. This was done to keep image sizes computationally reasonable (i.e., 320 x 320 pixels or smaller) and to allow for random image augmentation (e.g., flipping/mirroring, blurring/sharpening), which can improve the performance and generalization of neural networks (CITATION). The size of the windows depended on the resolution of the imagery created during the artificial coarsening procedure (see Table S1). One window was extracted for each digitized burrow present in each tile, along with five additional windows extracted for random locations within the tile. This process resulted in a total of 641 training images and 455 testing images.

All windowed image values were rescaled to the range 0 – 1 prior to training, which is known to improve the speed and accuracy of training deep neural networks (CITATION). The RGB image values were rescaled using the data range 0 – 255, the TPI values were rescaled using the data range -0.10 – 0.40, and the NDVI values already had a data range of 0 – 1, and thus did not require rescaling.

Image preparation was repeated at five spatial resolutions: 2, 5, 10, 15, and 30 cm. We resampled all image inputs to each resolution using the nearest neighbor technique. For the TPI, we first resampled the DEM and then recalculated TPI at the new resolution.

Training the deep neural network

Our neural network used the DeepLabV3+ architecture with a Resnet34 encoder initialized with pretrained weights using the imagenet dataset. We used the Mathew’s Correlation Coefficient loss function (Abhishek and Hamarneh 2021), a sigmoid activation function and stochastic gradient descent optimization with a triangular cyclic learning rate. The learning rate cycled between a base rate of 0.025 and an upper limit of 0.10, completing a full cycle back to the base rate every 4 epochs (one epoch is a complete iteration of all training images). We determined the learning rate range by running initial tests at a range of learning rates (1.0 × 10-5 to 0.5 × 10-1) with all inputs (RGB, TPI, NDVI) and setting the base learning rate to the value where model accuracy first increased and the upper learning rate to the value where accuracy leveled off (CITATION).

Out of the 641 training images, we randomly held out 20% (n=129) for training validation. We fed the remaining images (n=512) into the neural network in batches of 8. The images were randomly augmented using the albumentations package (CITATION). Augmentations included cropping, flipping, mirroring, blurring, motion blurring and sharpening (see Table S2). At the end of each learning rate cycle (i.e., every 4 epochs), the updated model was validated by predicting burrows in the held-out training data and calculating the F-score – our measure of accuracy – at the burrow scale, based on whether the centroid of the predicted burrow was within the digitized burrow polygon. The F-score can range between 0.0 and 1.0 and balances precision and recall, calculated as:

If the F-score improved by at least 0.005 over the previous learning rate cycle, we saved the model. We ran the training for a minimum of 2 learning rate cycles and a maximum of 16 learning cycles (i.e., 8 - 64 epochs), however if no improvement in F-score was observed for 2 consecutive learning rate cycles (i.e., 8 epochs), we stopped the training early and used the last saved model for all subsequent analysis.

Independent validation

We validated all the saved models at the scale of burrows and tiles using the independent testing data. At the burrow scale, we calculated precision, recall and the F-score based on whether the centroid of the predicted burrow was within the digitized burrow polygon. At the tile scale, we assessed the Person correlation and mean percent error (MPE) between predicted and observed burrow density.

**Colony classification and comparison to ground delineation**

We applied each model to the entire study area and extracted the centroid to generate a map of individual burrows. To scale the individual burrow predictions to delineate colonies, we used an approach similar to the ground-based method used to delineate colonies with handheld GPS units. First, we generated a 5-m grid over the study area and assigned to each grid cell the sum of burrows within a 25 m radius of each grid-cell center. Then, we created a binary grid of all cells with a value greater than 5, thus representing all locations in the study area with more than 5 burrows within 50 m of each other. To identify broad groupings of cells that meet this criteria, which we expect to represent active colonies, we adapted a method developed by Riitters et al. (2002) to classify forest connectivity. We calculated the density and connectivity of cells in the binary grid using a 11 x 11 cell (55 x 55 m) moving window. We then classified grid cells into five classes: Core (density ≥ 0.90), Perforated Core (density ≥ 0.60 and density ≥ connectivity), Edge (density ≥ 0.60 and connectivity > density), Transition (0.40 ≥ density < 0.60) and Non-colony (density < 0.40). Additionally, any grid cells completely surrounded by Core or Perforated Core were assigned to the Core classes, and any remaining grid cells completely surrounded by any combination of Core, Edge or Transition classes were assigned to the Transition class. We expected the Core classes to be the most likely to be part of an active colony, the Edge class to be uncertain and the Transition classes to be least likely.

We compared the model-derived colony classification to the ground-based colony delineation for each of the four pastures in the study area. This was done based on total hectares and the Jaccard score, which is calculated as the intersection of predicted and observed colonies divided by the union of the two (i.e., IOU score), and ranges between 0.0 (no overlap) and 1.0 (perfect overlap). We did this using the September imagery and ground data, from which the models were derived, and for the July imagery and ground data. The latter gave us an indication of how well a model might perform when vegetation conditions are different, in this case earlier in the growing season when vegetation is greener and cover and biomass are greater.

**Within-colony heterogeneity**

Colony age

We evaluated whether burrow density was related to the age of colonies using historical ground delineation of colonies from 2015 – 2021. For each grid cell within colonies delineated as active in 2021, we summed the number of consecutive years since 2015 the grid cell had been mapped as active. If the grid cell had not been delineated as active in 2021, but had been delineated as active in 2020, it was given a value of –1. If the grid cell had not been delineated in 2021 or 2020, it was assigned N/A (not active). We then created boxplots of model-predicted burrow density within each age class.

Vegetation

We evaluated whether burrow density was related to ground vegetation using satellite-derived estimates of standing biomass and bare ground cover. The satellite-derived estimates were created for a separate study using the Harmonized Landsat-Sentinel (HLS) dataset, produced at 30 m spatial resolution and interpolated at daily temporal resolution (see Kearney et al., 2022). We extracted the average burrow density within each 30 m HLS pixel and compared it to average biomass and bare ground values within 15 days of the UAV flights, as well as the 50-day change in biomass and bare ground leading up to the UAV flight. All biomass and bare ground values were converted to z-scores by subtracting the mean and dividing by the standard deviation of the region (i.e., the entire CPER area). We then created scatterplots of burrow density versus each metric, analyzed their linear relationships and plotted the mean and standard deviation within each predicted colony class (i.e., Core, Edge, Transitional and Not active).

**RESULTS**

**Burrow detection**

Original (2 cm) resolution

At the finest resolution of 2 cm, training validation performed well (burrow-scale F-score 0.84 - 0.93) for all the input combinations we tested, except when using only NDVI (F-score 0.55; Table S3). We therefore did not consider the NDVI-only model in further analysis. At the 2 cm resolution, the model using only TPI performed best (F-score 0.87) and including additional inputs along with TPI did not improve the F-score (Table 1). Performance was notably lower in models that did not include TPI; the RGB and RGB + NDVI models had F-scores of 0.73 and 0.66, respectively.

The ground dataset of geolocated burrows (n=273) showed that, at 2 cm resolution, omitted burrows (false negatives) tended to be smaller, less active and more vegetated compared to detected burrows (true positives) regardless of the inputs used (Figure 2). The detected burrows were marginally larger, more active and less vegetated than average, and this trend was slightly more pronounced for models that did not include TPI. The ground dataset of non-burrow features (n=108) showed that commission error (false positives) of anthills was lowest for TPI, however the models with both RGB and NDVI tended to have lower commission error for most, but not all, of the other burrow-like features (Table 2).

Burrow densities predicted by models were strongly and significantly correlated to the density of manually digitized burrows for the 30 x 30 m validation tiles (Figure 3). Again, models that included TPI performed best, with correlation coefficients of 0.94 - 0.95 and low bias (MPE = -0.01 - 0.13). The RGB model had a correlation coefficient of 0.88 and MPE of 0.03. The RGB + NDVI model had a strong correlation (r = 0.84) but tended to overpredict burrow density (MPE = 0.46).

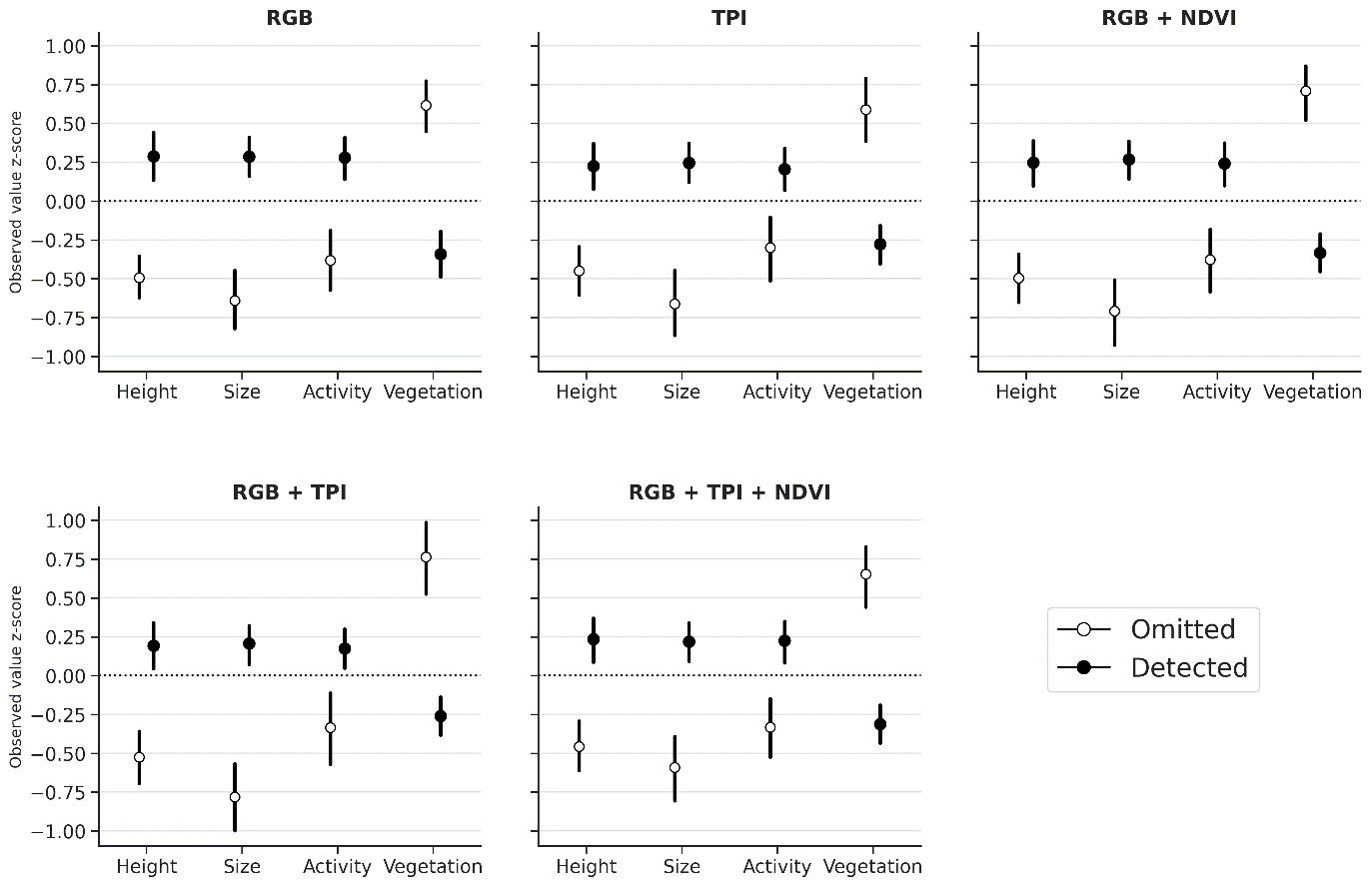
We were able to compare the validation of predictions to a validation of the manual digitization for a small subset of ground geolocated burrows (n=47) that overlapped the manually digitized 30 x 30 m tiles (n=13). We found that accuracy of manual digitization was high, and that 2 cm models with TPI slightly outperformed manual digitization. For the 47 burrows, manual digitization achieved an F-score of 0.77, while the predictions from models that included TPI had F-scores of 0.78 - 0.82, and models without TPI had F-scores of 0.67 - 0.69 (Table S4). The correlation between burrow density derived from manual digitization vs ground geolocation for the 13 tiles was 0.93 (Figure S1), whereas for the model predictions vs. ground geolocation, correlation ranged from a low of 0.84 for the RGB model to 0.94 for the RGB + TPI model (Figure S2).

**Table 1.** Independent burrow-scale validation of 2 cm resolution models using the test images derived from the 20% hold-out tiles and tiles covering ground data (n=455 images).

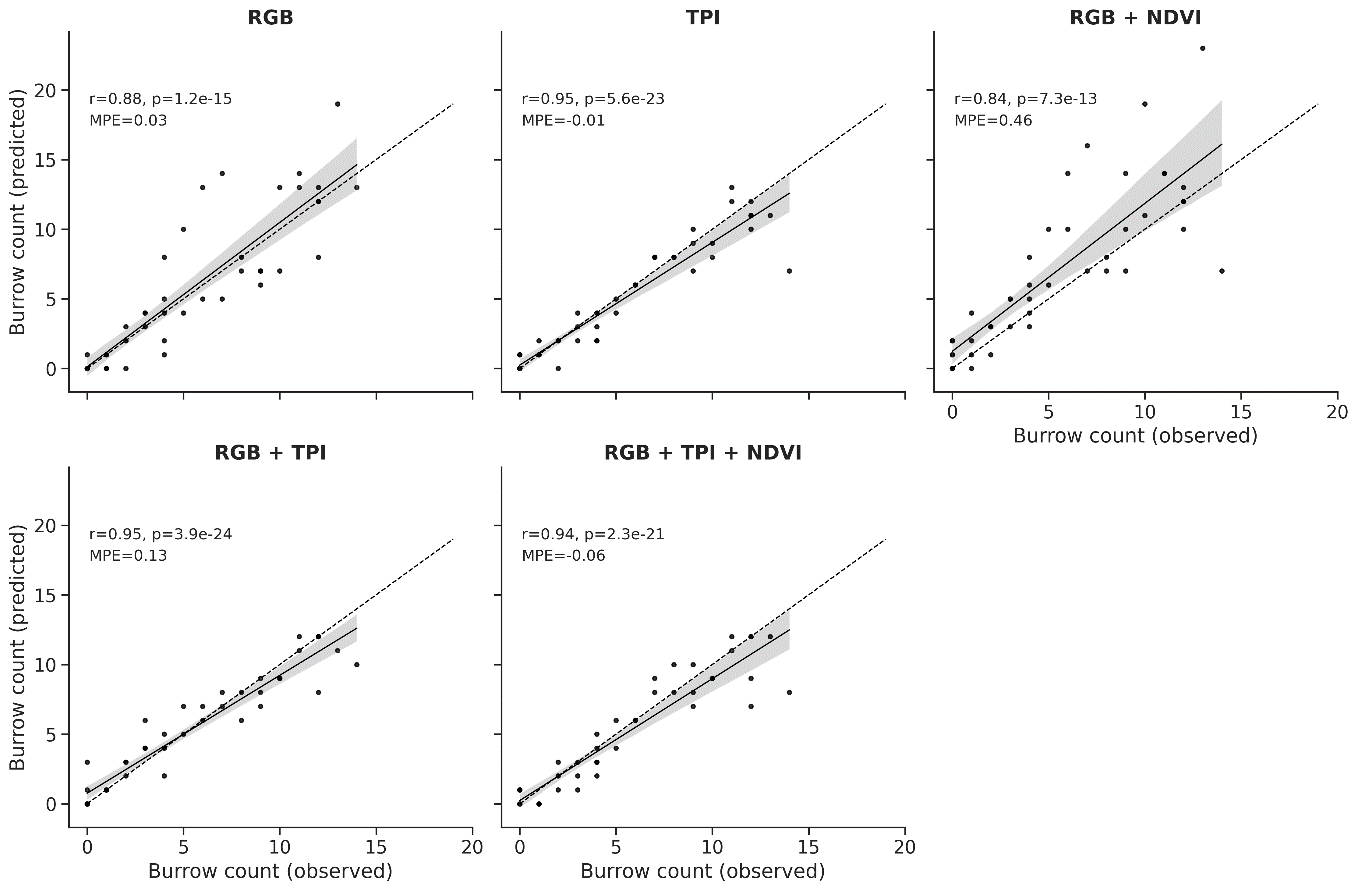
|  |  |  |  |
| --- | --- | --- | --- |
| **Inputs** | **Precision** | **Recall** | **F-score** |
| TPI | 0.90 | 0.84 | 0.87 |
| RGB + TPI | 0.86 | 0.84 | 0.85 |
| RGB + TPI + NDVI | 0.88 | 0.81 | 0.84 |
| RGB | 0.70 | 0.75 | 0.73 |
| RGB + NDVI | 0.59 | 0.76 | 0.66 |

**Table 2.** Commission error (false-positive) rates of 2 cm resolution models for ground-mapped features expected to appear similar to active burrows.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Feature type** | **n** | **RGB** | **TPI** | **RGB + NDVI** | **RGB +**  **TPI** | **RGB +**  **TPI + NDVI** |
| Anthill | 67 | 0.07 | 0.01 | 0.18 | 0.03 | 0.03 |
| Den | 8 | 0.38 | 0.50 | 0.12 | 0.50 | 0.12 |
| Digging | 6 | 0.33 | 0.17 | 0.17 | 0.17 | 0.17 |
| Old burrow | 27 | 0.07 | 0.11 | 0.07 | 0.11 | 0.15 |
| *Overall* | 108 | 0.06 | 0.05 | 0.08 | 0.05 | 0.04 |



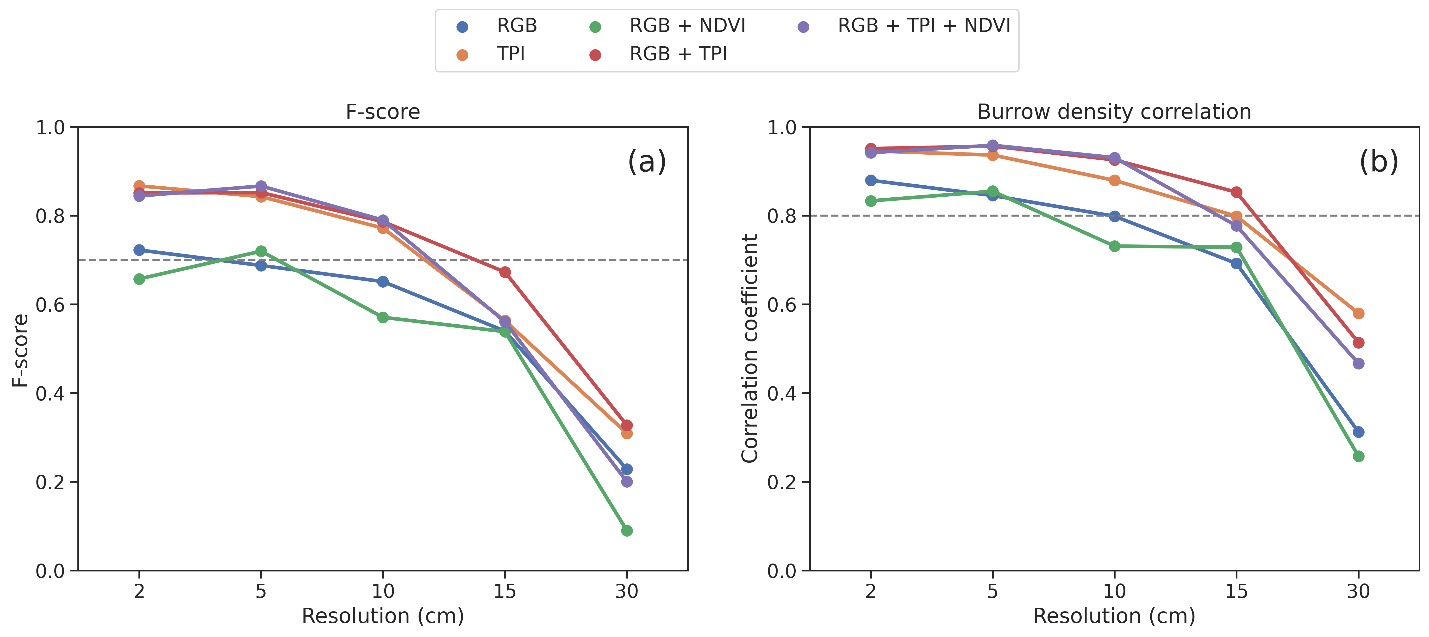
**Figure 2.** Patterns of omission error (false negatives) for 2 cm resolution models compared to detected burrows (true positives) for ground-mapped active burrows (n=273). Y-axis represents the z-score (standard deviations away from the mean) of the group compared to the entire dataset.



**Figure 3.** Correlation between model predicted and manually digitized (observed) burrow densities at the scale of 30 x 30 m testing tiles (n = 45). ‘r’ is the Pearson correlation coefficient, ‘p’ is the alpha (significance) of the correlation and ‘MPE’ is the mean percent error of predictions, an indication of bias.

Coarser resolutions

When imagery was coarsened to 5 cm, test validation performance was little changed compared to 2 cm; it improved slightly for models that included NDVI as in input, remained the same for models with TPI as an input and declined slightly for the RGB only model (Figure 4; Table S5). In general, performance began to decline slightly at 10 cm resolution and decreased more sharply when imagery was coarsened to 15 and 30 cm, though the rate of decline varied across models. At 10 cm, the three models with TPI continued to perform best with F-scores of 0.77 - 0.79 (Figure 4a) and burrow density correlations of 0.88 - 0.93 (Figure4b). The F-scores and burrow density correlations of the RGB and RGB + NDVI models fell below 0.70 and 0.80, respectively, at 10 cm. At 15 cm the RGB + TPI model performed best (F-score = 0.67). It was the only model with a burrow density correlation greater than 0.80 (r = 0.85) and recall greater than 0.50 (recall = 0.53). Precision remained moderate or high for all models at 15 cm (0.68 - 0.92), indicating that, up to 15 cm image resolution, models not spurious, rather they were simply omitting more burrows. However, at 30 m resolution, performance degraded for all models as a result of both poor recall and poor precision (Table S5), indicating more spurious performance.

**Figure 4.** Model test validation at increasingly coarse spatial resolution of imagery for each combination of model inputs. F-score (a) is calculated from individual burrows across all validation images and burrow density correlation (b) is calculated from the 30 x 30 m validation tiles.

**Colony delineation**

Prediction for season of model training (September)

We focused our colony-scale evaluation on models using the 5, 10 and 15 cm imagery. This was done for two reasons. First, predicting burrows across the entire study using the 2 cm image inputs proved computationally challenging, largely due to the processing time required to generate the TPI layer. Second, since burrow detection of the 30 cm imagery was poor, we did not consider it further at the colony scale. We also dropped the RGB + NDVI model since burrow detection performance tended to be poorer compared to the RGB only model.

The areas predicted as Core colony from the 5 cm models had moderate to high overlap with the ground delineation of active colony for the three pastures with large active colonies in September (22EW, CN, 5W), and overlap generally varied more by pasture than by model (Figure S4). Of the three active pastures, the highest Jaccard scores were in pasture 22EW (0.89 - 0.94) and the lowest in pasture 5W (0.58 - 0.61; Figure S4). For pasture CN, the three models with TPI had Jaccard scores of 0.82 - 0.85 and the RGB model had a score of 0.55 (Figure S4). For these three pastures combined, the RGB model had the closest predicted total Core colony area compared to the total ground delineated active colonies. However, this was due to relatively large offsetting differences in different colonies (Table S6), and when we looked at Jaccard score and the relative absolute difference (which accounts for colony area), we found that RGB + TPI and RGB + TPI + NDVI models performed best across the active pastures, with pasture-scale absolute differences relative to the ground delineation of around 11%, compared to a difference of around 15% for the TPI model and 24% for the RGB model (Table S7).

Pasture 29-30, in which colonies experienced a plague epizootic during the study, had only 0.70 ha delineated on the ground as active colony in September. Overlap between model-predicted Core colony and ground delineated active colony was essentially zero for all 5 cm models, which predicted between 37.25 ha (RGB) and 120 ha (TPI) of Core colony (Table S6). In this case, models that included TPI had lower performance due to detection of recently uninhabited burrows. Figure 5 shows that much of the area predicted as Core colony by the RGB + TPI model corresponds to areas delineated as active in September 2020 but contracted due to plague throughout 2021. We note that results were similar for the other models. When evaluated across all four pastures, absolute differences relative to the ground delineation were: RGB (35%), RGB + TPI (37%), RGB + TPI + NDVI (40%), TPI (50%) (Table S7).

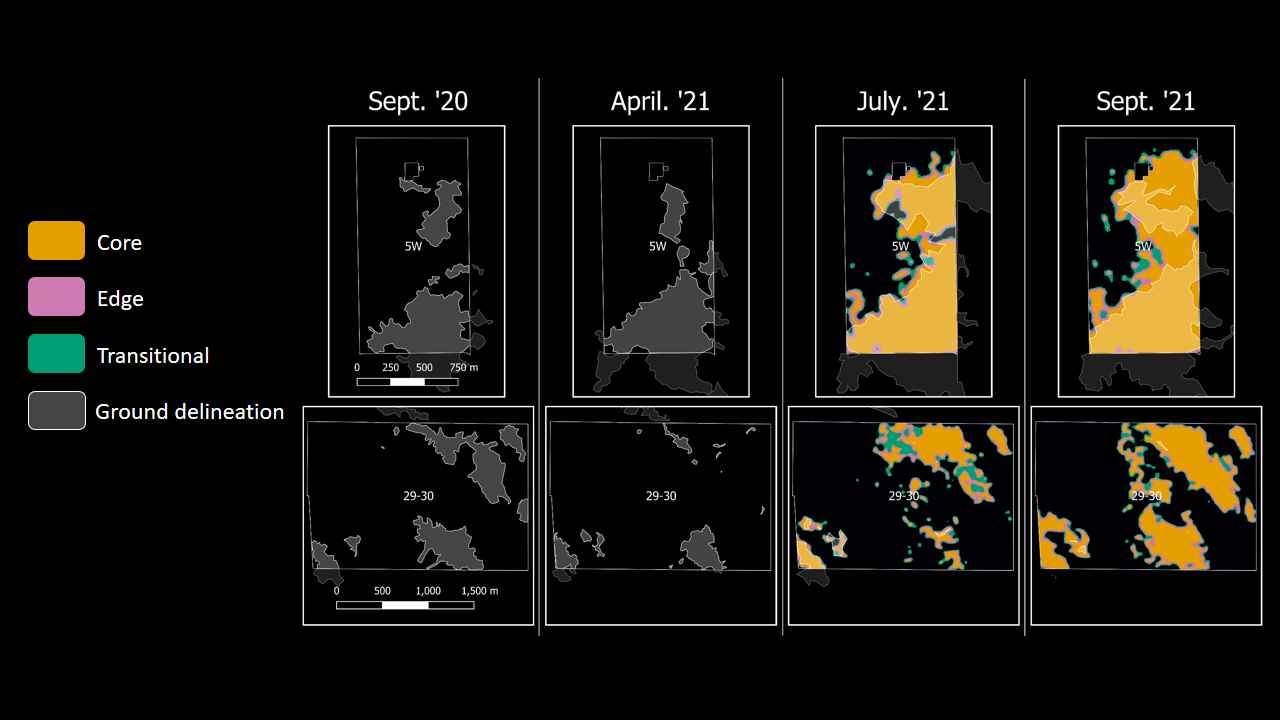
At 10 cm resolution, Jaccard scores only changed marginally for most models and pastures, with some increasing slightly, some decreasing slightly and some remaining essentially unchanged (Figure S3). However, the area predicted as Core colony in pasture 29-30 was much closer to zero for the three models that included TPI (Table S6). This improvement, combined with only a marginal reduction in performance for the active pastures, resulted in an improvement of the overall Jaccard score when evaluated across all pastures for the three models with TPI (Table S6). At 15 cm, Jaccard scores decreased for most models and pastures, although less-so for the TPI only and RGB + TPI models (Figure S3).

Prediction in a new season (July)

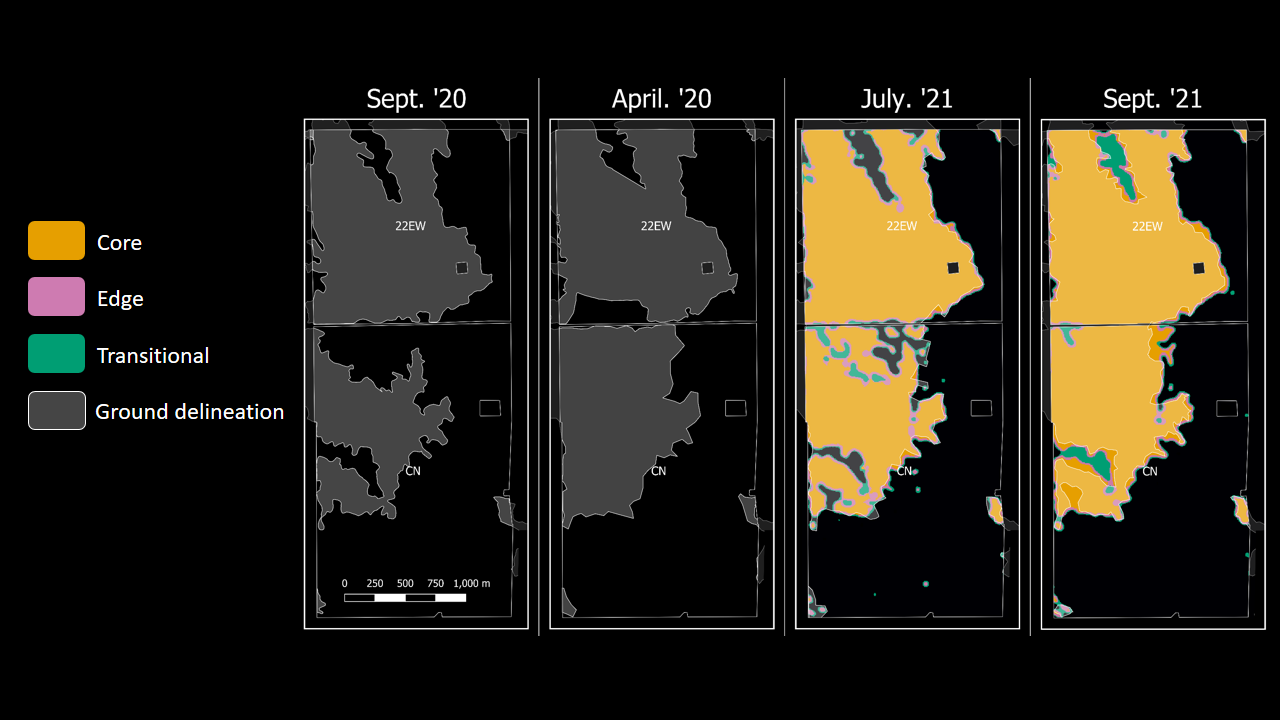
When evaluated for July, a season for which the model was not trained, overlap of model-predicted Core colony and ground delineated active colony were similar to what we observed for September. All the 5 cm models had moderate to high overlap (0.66 - 0.88) with the ground delineation of active colony for the three pastures with large active colonies, except for the RGB model in pasture CN (0.25), which predicted substantially less area as Core colony compared to ground delineation (Figure S3). As with September, the models that included TPI tended to have higher overlap than the RGB model for July. The three models with TPI had absolute relative differences in area of 11 – 18 % compared to the ground delineation in the active pastures, while the RGB model had a difference of 38% (Table S7).

Overlap was slightly higher in pasture 29-30 (0.08 - 0.22) for the July predictions compared to September. This resulted in a slightly lower overall absolute relative difference in colony area across all pastures, ranging from 20% for the RBG + TPI + NDVI model to 39% for the RGB model (Table S7).

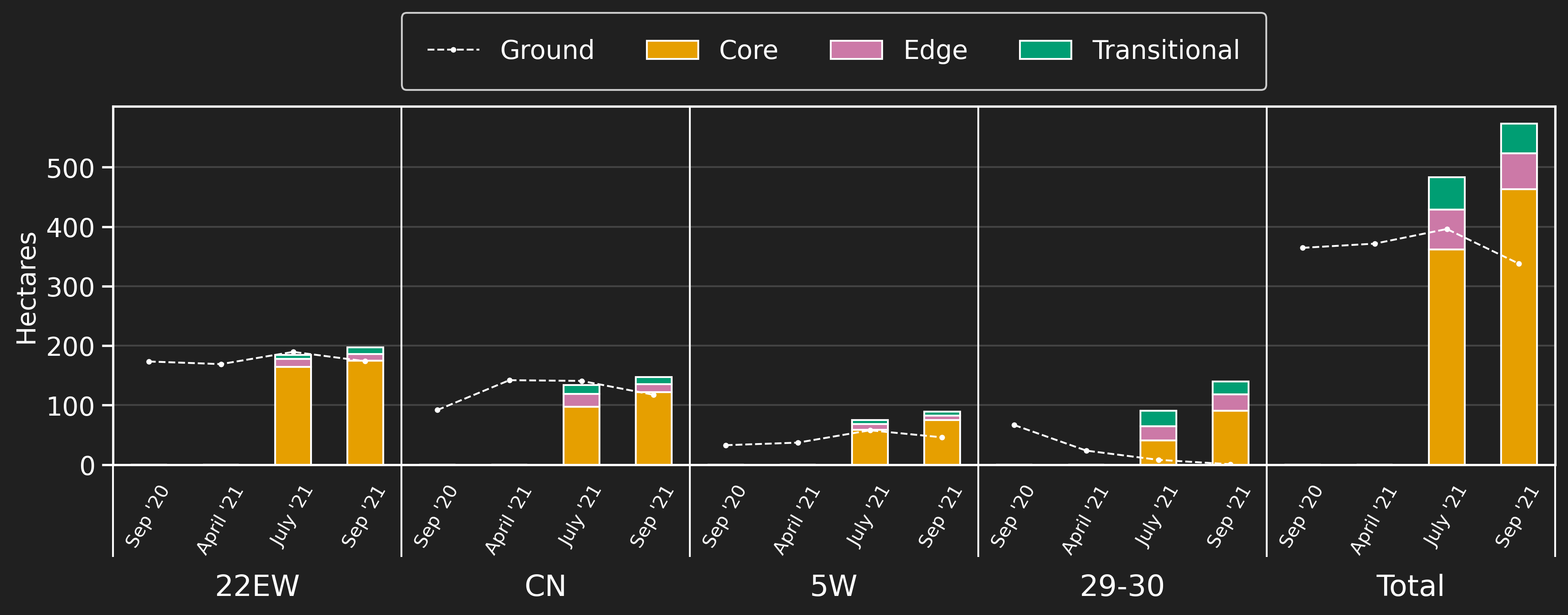
At coarser spatial resolutions, Jaccard scores dropped off more quickly for July compared to September, and they were below 0.70 in all pastures (Figure S3). At 10 cm, overall absolute relative differences in area were 45 – 88%, and at 15 cm they were 46 – 97% (Table S7).



**Figure 5.** Ground delineation (grey with white outline) from all colony surveys in 2020 and 2021 for pastures 5W and 29-30 compared with predicted colony classes from the 5-cm RGB + TPI model for July and September of 2021. Note that these pastures are not spatial adjacent and are each shown at different scales (scale bar in left-most panel for each pasture).



**Figure 6.** Ground delineation (grey with white outline) from all colony surveys in 2020 and 2021 for pastures 22EW and CN compared with predicted colony classes from the 5-cm RGB + TPI model for July and September of 2021. Note that these pastures are spatial adjacent, separated by a gravel road.



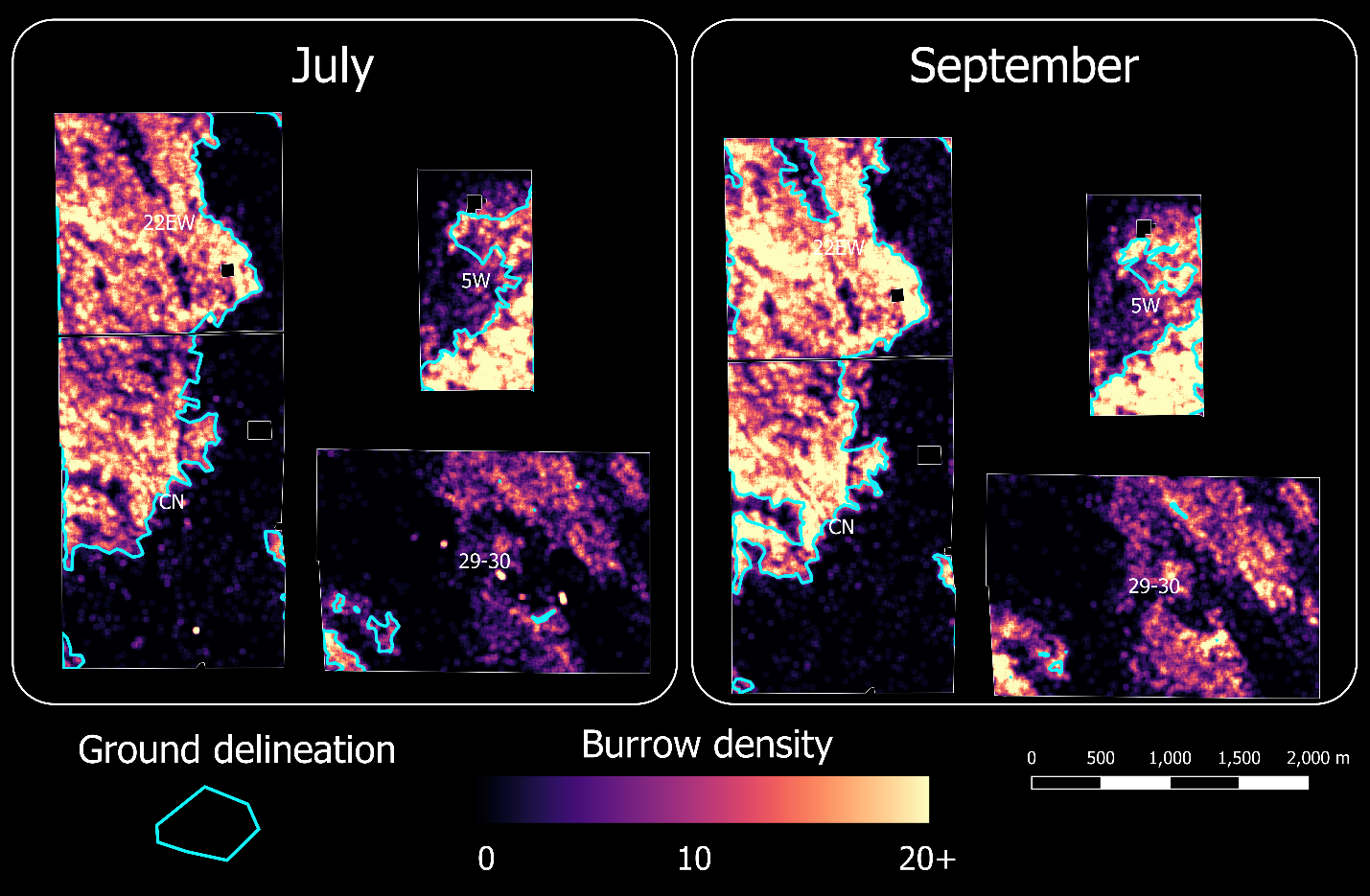
**Figure 7.** Trends in 2020-2021 ground delineated area of active prairie dog colony (dotted lines) compared to the model-predicted area in the three potential colony classes (bars) for July and September of 2021, the two months for which UAS imagery was available.

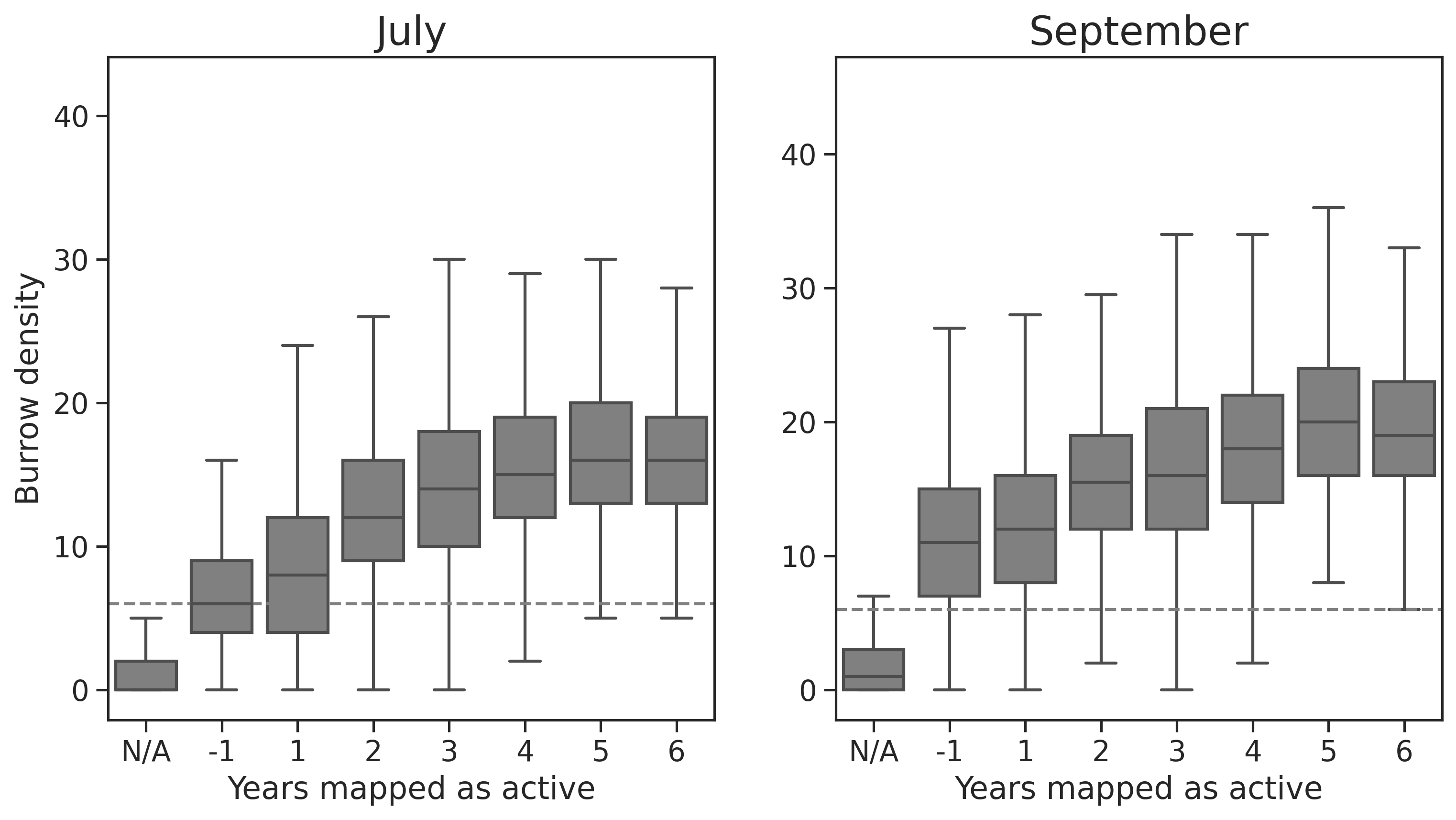
**Burrow density and colony heterogeneity**

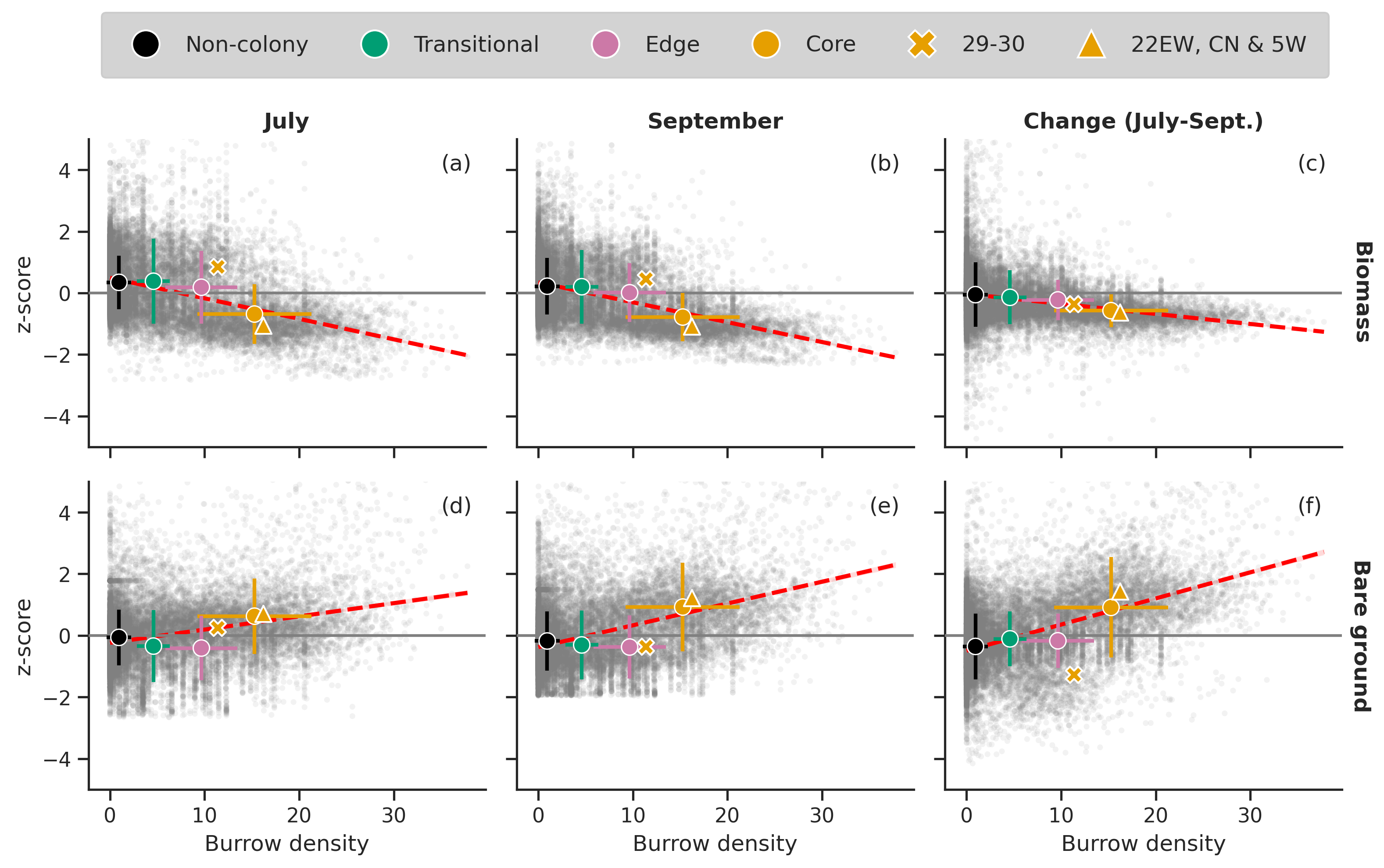
To assess heterogeneity of burrow density and its relationship to colony age and vegetation metrics, we used burrow density predictions from the 5 cm RGB + TPI model, shown in Figure 8. We chose this model since it performed well in our burrow-scale evaluation for September and at the colony scale for both seasons.

For a given age class, model-predicted burrow density was generally higher for September compared to July (Figure 9). Burrow density tended to increase with colony age for both July and September. There was a clear difference in burrow density between non-active areas within the study pastures (“N/A” in Figure 9) and areas delineated as active in either 2020 or 2021. There was little difference in burrow density between the areas delineated as active for just one year compared to areas delineated as active in 2020 but not in 2021 (i.e., colony age of –1), especially for September. However, there was a clearer difference between the areas with a colony age of -1 and areas delineated as active for two or more consecutive years (Figure 9).

Burrow density was negatively correlated with standardized biomass for July (Pearson r = -0.43) and September (r = -0.46) and was positively correlated with bare ground cover for both seasons, respectively (r = 0.27 and r = 0.41; Figure 10). Burrow density was also negatively correlated with the change in biomass between July and September (r = -0.27) and positively correlated with the change in bare ground cover (r = 0.46). Across all pastures, areas predicted as Core colony had higher burrow density, lower biomass and higher bare ground cover compared to areas not predicted to be part of a colony (Figure 10). Areas predicted as Edge or Transitional had higher burrow density compared to non-colony areas, but biomass and bare ground cover were similar to the non-colony area and to the overall mean of the region. The area predicted as Core colony in pasture 29-30, which experienced plague during the 2020-2021 period, had lower burrow density, higher biomass and less bare ground cover compared to the Core colony area for the other three pastures. The change in bare ground cover from July to September was markedly lower for pasture 29-30 compared to the other three pastures and compared to the mean of the region (Figure 10f). It was also much lower than would be expected based on the linear trend between burrow density and July-September bare ground cover change.

**Figure 8.** Maps of model-predicted prairie dog burrow density (heatmap, warmer colors indicate higher density) overlayed with ground delineated boundary of active colonies (light blue polygons). Note that pastures are not actually spatially adjacent (see Figure 1) but are all shown at the same scale.

**Figure 9.** Variation in September 2021 model-predicted prairie dog burrow density (within 25 m radius) compared to the number of years an area has been mapped as an active colony using ground delineation. “N/A“ refers to an area that was not mapped as active in 2020 or 2021, and ”-1” refers to areas mapped as active in 2020 but not in 2021. The dotted line shows a burrow density of 6, the threshold used in the model-predicted colony delineation procedure.



**Figure 10.** Scatterplots of burrow density (within a 25 m radius) and satellite-derived metrics for July, September and the slope of the change in metrics over the July-September period. Grey dots represent individual 30 m pixels for the entire Central Plains Experimental Range (CPER) and are z-score standardized by subtracting the mean and dividing by the standard deviation of the entire CPER. Red dotted lines show the linear relationship for all CPER for each variable pair. Means and standard deviations are shown for the four model prediction classes (Non-colony, Transitional, Edge, Core) within the four study pastures. Means are shown for the Core areas in pasture 29-30, which experienced plague in 2020/21 (‘X’ symbol), and for the other three pastures (22EW, CN, 5W) with large, active colonies (‘Δ’ symbol).

**DISCUSSION**

**Burrow detection and colony classification accuracy**

Overall, agreement between burrow-scale and colony-scale model predictions and the corresponding validation datasets - burrow digitization, ground georeferencing of burrows and ground delineation of colonies - were high. However, it is important to note that each of the validation datasets have their own sources of uncertainty and no one dataset can be considered perfect truth.

At the burrow scale, our results show that, for this study area, we can expect around 80 – 85% harmonic accuracy (F-score) of burrow detection using a deep CNN model trained with a moderately sized digitization dataset from that same imagery. In fact, our results suggest that models may even slightly outperform manual digitization, even when trained with that same imperfectly digitized data (e.g., Table S4, Figure S1-S2). There are limited other studies that have attempted to detect individual prairie burrows from remotely sensed imagery, and, to our knowledge, this is the first attempt to use a CNN for this purpose. Pixel-based approaches have generally had very poor accuracy due to the strong spectral similarity of burrows to other land cover typified by bare ground cover (e.g., Folluo 2019; Jost et al., 2018). A few studies have achieved results using object-based image analysis (e.g., Delparte et al., 2019; Folluo 2019), however these approaches tend to require licensed software and rely on computationally expensive object detection followed by developing a classification ruleset, which can be subjective (if manually defined) and have limited transferability.

At the colony scale, we observed 60 – 94% overlap in September between model predictions and ground mapping for the three pastures with active colonies. It was encouraging that overlap remained high for the July predictions and ground mapping (68 – 86%) for these three pastures. This suggests that well-trained models may be robust across seasons and changing vegetation conditions for active colonies. However, further testing across different soil types, plant communities and vegetation conditions (e.g., drought) are needed to determine model transferability over space and time. Additionally, it is possible that more robust techniques to scale from individual burrow detection to colony delineation can be developed.

Identifying recent drops in colony activity (e.g., from plague, lethal control) may prove challenging. Ground assessment of colony activity relies on a variety of indicators beyond burrow presence that are likely impractical to identify from remotely sensed imagery, such as presence of prairie dogs, feces, tracks, digging, cobwebs in burrow entrances and more. Recently uninhabited burrows are likely to persist on the landscape for longer than these other indicators after colony activity ceases. In this study, we were only able to assess results for a plague event that occurred within the same growing season as image acquisition. There was widespread false-positive detection of colony areas that had become inactive in the last few months, especially using the September imagery. Models also predicted a larger Core colony area using the September imagery compared to July, despite ground mapping showing a contraction of the colony between July and September. This suggests that late-season imagery may be more like to detect recently inactive burrows (e.g., due to reduced vegetation cover later in the season in this system) and that image acquisition during peak vegetation production may produce better results for colonies that have recently become inactive.

Using near-real time satellite monitoring of vegetation conditions may also help to identify recently uninhabited colonies. We found that the area in pasture 29-30 identified as Core colony by our model had higher biomass, less bare ground cover and a much lower slope of change in bare ground cover than expected given the burrow density of that area (Figure 10). Typifying the relationship between burrow density, vegetation metrics and known active colonies for a given region may enable early detection of colony abandonment, even when burrows persist on the landscape. More research is needed to better understand how long it takes for uninhabited burrows to no longer be detected by the neural network, as well as whether robust methods can be developed to flag recently uninhabited colonies using satellite imagery.

**Burrow density and colony heterogeneity**

It is clear from the September imagery that predicted burrow density is strongly correlated with burrow density derived from manual digitization (Figure 3) and observed on the ground (Figure S2). It is less how accurate burrow density predictions were when the model was applied to July imagery, since we did not perform burrow-scale validation for July. We observed an increase in predicted burrow density between July and September, which could be result of (a) new burrow construction, (b) higher detection rates of burrows later in the season when vegetation cover is declining and/or (c) poorer detection accuracy in July since the model was trained for September conditions (e.g., with less green vegetation). If new burrows are being constructed at a rate faster than uninhabited burrows are demolished (i.e., (a) above), or if we expect to detect more inactive burrows when vegetation cover is low (i.e., (b) above; also see previous section), this could make monitoring active burrow density change or active colony expansion and contraction challenging over short periods of time, for example within a single growing season.

The fact that burrow density increased in Pasture 29-30, in which colony activity was substantially reduced by plague, suggests that causes (b) and (c) noted above are at least contributing factors. Additional testing using data from this study, for example developing a new model for July or with July and September imagery combined – would help to better understand the relative contributions of (b) vs (c). Repeating this study across multiple years and seasons to capture a wider range of vegetation conditions would help to answer this question more definitively.

Creating a much larger training dataset may also improve model accuracy and consistency. Manual digitization is relatively fast and straightforward, provided users have access to a custom application such as that used in our study, or have access and training for GIS software. Based on our study, we expect the accuracy of manual digitization to be high (e.g., Table S4 and Figure S1) and consistent across users, provided they have some training and/or knowledge of the area. Other studies have also indicated that manual digitization of burrows from imagery corresponds well with ground data (e.g., Delparte et al., 2019).

Maps of burrow density will open many new monitoring and research opportunities. Burrow density has been shown to be strongly correlated with population density (Johnson and Collinge, 2004). Maps of burrow density could help map habitat for associated species that prey on prairie dogs or be integrated into models of colony dynamics. For example, Cincotta et al. 1988 showed that burrow density at the edge of a colony influenced the probability that expansion would occur. Other studies suggest that colonies with higher prairie dog density may have higher plague transmittance rates compared to less dense colonies ([Cully and Williams, 2001;](https://www.sciencedirect.com/science/article/pii/S0006320703001654#BIB17) Others?) and burrow density maps could become inputs for models forecasting plague (Davidson et al, 2022) and subsequent recolonization from nearby colonies. Our study shows that burrow density is correlated vegetation conditions. Greater burrow density may be expected to enhance habitat for some species (e.g., Mountain Plover, Burrowing Owls), diminish habitat for other species (e.g., Grasshopper Sparrow) and reduce immediate forage availability for domesticated livestock (Augustine and Derner 2021..). Network and clustering analyses of burrow configuration could help to identify individual coteries and better understand spatial organization within colonies (Hasan 2019).

**Key takeaways for future operational monitoring**

Our results suggest that having RGB and TPI at spatial resolutions of 5 – 15 cm will be ideal for future monitoring; resolutions finer than 5 cm present computational challenges and resolutions greater than 15 cm likely fail to resolve individual burrows. Adding NDVI generally did not improve results at the burrow or colony scale for any spatial resolution, and in some cases decreased performance. The TPI was the most important input for consistent burrow detection, likely in part due to reduced false-positives for anthills and other features typified by circular bare ground patches, but lacking the distinctive burrow mound and entrance feature highlighted by the TPI. Combining RGB with TPI performed well at the colony scale (Table S6) and improved burrow detection at coarser image resolutions greater than 10 cm (Table S5, Figure 4), at which point TPI images likely lose their ability to resolve smaller burrow entrances. Additionally, using only TPI as an input may increase the likelihood of detecting recently uninhabited burrows, leading to over-estimation of colony areas in recently uninhabited colonies (Table S7). It is possible that the RGB only model may do better with recently uninhabited colonies. For example, in both seasons, the RGB model predicted a much smaller area of Core colony – more similar to the ground delineation – in the recently plagued pasture 29-30, without much change in the predicted area for two of the other three pastures. In pasture CN, the predicted Core area was reduced compared to the RGB + TPI model, however most of the change was to the interior of the colony, with little change in the perimeter. The ground delineation approach does not involve intensive surveys of the interior of colonies, and it is possible that those interior areas do in fact have less or no activity.

It is important to note that our original 6-cm DEM, which we resampled to finer or coarser resolutions before creating the TPI’s in our study, was initially derived from a point cloud with a density of around 220 points m-2 created photogrammetrically from the 1.5 - 1.9 cm RGB imagery. While we found that models using the coarsened 10 – 15 cm resolution TPI still performed well, it not clear whether a DEM originally derived from 10 – 15 cm RGB imagery or from airborne laser scanning (a.k.a. LiDAR) would perform equally well. In our tests, we resampled the DEM at a pixel scale using a simple box average, but did not resample the original point cloud. Based on visual inspection in a GIS, burrow entrance depressions typically had a diameter of 15 - 40 cm. Using an equation proposed by Cekada et al. (2010), the smallest theoretical point density required to resolve features of this size range from 25 pts/m^2 (for 40 cm graphical accuracy) to 178 pts/m^2 (for 15 cm graphical accuracy). Further testing is needed to better understand whether TPI derived from point clouds in this range perform as well as our artificially coarsened TPI, especially at the lower end of the range which could be achieved with LiDAR. With current technology, creating point clouds at the upper end of this range may be cost-prohibitive over large areas. It is also possible that, given a sufficiently large training dataset, adequate accuracy could be achieved using only RGB imagery, though our results suggest TPI will generally improve results.

The fixed-wing Trinity F90+ UAS was critical for acquiring imagery across the entire 1,120-ha study area in a timely manner. However, the platform has limited capabilities to fly in windy conditions, and could not be flown at windspeeds greater than 16-20 kph (4.5 ~ 5.5 m sec-1). North American prairie dog colonies tend to be in some of the windiest parts of the continent, which could limit the ability to use this platform for regular monitoring of large areas. Manned aircraft may be required to monitor vast areas of windy and rugged terrain occupied by prairie dogs. Sharing costs with other monitoring programs (e.g., vegetation, critical habitat, fire fuels) may be necessary to enable widespread image acquisition at the necessary resolution.

Image-based prairie dog burrow detection and colony delineation is one potential new tool for monitoring, but likely it will be best utilized in conjunction with other current and new monitoring techniques. Maps of burrow density, colony prediction class (e.g., Core, Edge and Transition) and satellite-derived vegetation metrics could be used to identify areas of high uncertainty for colony delineation and these areas could be targeted for efficient ground surveys. Some level of ongoing ground validation may also help to build confidence in models, improve models with new training data or identify limitations.

**CONCLUSIONS**

**SUPPLEMENTARY MATERIALS**

**Supplementary Tables**

**Table S1.** Range of window sizes used at each spatial resolution to extract training images from within the 30 x 30 m training tiles. Image extraction for each spatial resolution was performed by randomly selecting the window size at any 32-pixel interval within the specified range.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Window size** | | | |
|  | **Minimum** | | **Maximum** | |
| **Resolution (cm)** | *Pixels* | *Meters* | *Pixels* | *Meters* |
| 2 | 224 x 224 | 4.5 x 4.5 | 320 x 320 | 6.4 x 6.4 |
| 5 | 96 x 96 | 4.8 x 4.8 | 192 x 192 | 9.6 x 9.6 |
| 10 | 64 x 64 | 6.4 x 6.4 | 128 x 128 | 12.8 x 12.8 |
| 15 | 32 x 32 | 4.8 x 4.8 | 96 x 96 | 14.4 x 14.4 |
| 30 | 32 x 32 | 9.6 x 9.6 | 64 x 64 | 19.2 x 19.2 |

**Table S2. Image augmentations and their probabilities.**

|  |  |
| --- | --- |
| **Augmentation(s)** | **Probability** |
| Crop within range (see Table S1) | 1.0 |
| Horizontal flip | 0.5 |
| Vertical flip | 0.5 |
| One of:  Sharpen  Blur  Motion blur | 0.25 |

**Table S3.** Model training validation results at the burrow and pixel scales for all spatial resolutions. F-score = F1 score, calculated as the harmonic mean of precision and recall; MCC = Mathew’s Correlation Coefficient; IOU = Intersection Over Union, also called Jaccard’s score or J-score

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  | **Pixel scale** | | | | | |
|  | **Inputs** | **F-score (Burrow)** | **Best epoch** | **MCC loss** | **IOU score** | **F-score** | **Precision** | **Recall** | **Accuracy** |
| 2 cm | RGB | 0.86 | 16 | 0.79 | 0.75 | 0.86 | 0.89 | 0.84 | 0.99 |
| 2 cm | NDVI | 0.55 | 16 | 0.94 | 0.52 | 0.72 | 0.76 | 0.69 | 0.99 |
| 2 cm | TPI | 0.89 | 16 | 0.78 | 0.77 | 0.88 | 0.90 | 0.87 | 0.99 |
| 2 cm | RGB + NDVI | 0.84 | 12 | 0.81 | 0.74 | 0.86 | 0.86 | 0.85 | 0.99 |
| 2 cm | RGB + TPI | 0.93 | 12 | 0.75 | 0.81 | 0.90 | 0.92 | 0.88 | 0.99 |
| 2 cm | RGB + TPI + NDVI | 0.93 | 12 | 0.76 | 0.81 | 0.90 | 0.89 | 0.91 | 0.99 |
| 5 cm | RGB | 0.74 | 8 | 0.81 | 0.66 | 0.80 | 0.91 | 0.71 | 1.00 |
| 5 cm | TPI | 0.81 | 20 | 0.74 | 0.71 | 0.83 | 0.87 | 0.80 | 1.00 |
| 5 cm | RGB + NDVI | 0.71 | 12 | 0.85 | 0.64 | 0.79 | 0.92 | 0.69 | 0.99 |
| 5 cm | RGB + TPI | 0.85 | 20 | 0.72 | 0.73 | 0.85 | 0.88 | 0.83 | 1.00 |
| 5 cm | RGB + TPI + NDVI | 0.87 | 16 | 0.72 | 0.75 | 0.86 | 0.89 | 0.84 | 1.00 |
| 10 cm | RGB | 0.65 | 12 | 0.75 | 0.55 | 0.73 | 0.71 | 0.75 | 1.00 |
| 10 cm | TPI | 0.78 | 8 | 0.74 | 0.65 | 0.80 | 0.87 | 0.75 | 1.00 |
| 10 cm | RGB + NDVI | 0.67 | 16 | 0.84 | 0.58 | 0.74 | 0.92 | 0.62 | 1.00 |
| 10 cm | RGB + TPI | 0.83 | 28 | 0.69 | 0.71 | 0.84 | 0.91 | 0.78 | 1.00 |
| 10 cm | RGB + TPI + NDVI | 0.81 | 16 | 0.68 | 0.69 | 0.83 | 0.88 | 0.78 | 1.00 |
| 15 cm | RGB | 0.68 | 28 | 0.75 | 0.59 | 0.76 | 0.82 | 0.71 | 1.00 |
| 15 cm | TPI | 0.59 | 16 | 0.77 | 0.51 | 0.71 | 0.69 | 0.72 | 1.00 |
| 15 cm | RGB + NDVI | 0.63 | 24 | 0.79 | 0.55 | 0.73 | 0.80 | 0.67 | 1.00 |
| 15 cm | RGB + TPI | 0.73 | 20 | 0.75 | 0.61 | 0.77 | 0.87 | 0.70 | 1.00 |
| 15 cm | RGB + TPI + NDVI | 0.62 | 36 | 0.79 | 0.52 | 0.72 | 0.78 | 0.67 | 1.00 |
| 30 cm | RGB | 0.61 | 64 | 0.77 | 0.54 | 0.70 | 0.75 | 0.66 | 1.00 |
| 30 cm | TPI | 0.23 | 8 | 0.88 | 0.17 | 0.39 | 0.28 | 0.62 | 0.99 |
| 30 cm | RGB + NDVI | 0.46 | 24 | 0.95 | 0.44 | 0.62 | 0.94 | 0.46 | 1.00 |
| 30 cm | RGB + TPI | 0.58 | 28 | 0.80 | 0.56 | 0.71 | 0.85 | 0.61 | 1.00 |
| 30 cm | RGB + TPI + NDVI | 0.47 | 8 | 0.85 | 0.46 | 0.63 | 0.89 | 0.49 | 1.00 |

**Table S4.** Burrow-scale validation for manual digitization and 2 cm model predictions for ground geolocated burrows within tiles that were also digitized (n=47 burrows; n=13 tiles).

|  |  |  |  |
| --- | --- | --- | --- |
| **Inputs** | **Precision** | **Recall** | **F-score** |
| *Manual digitization* | *0.78* | *0.77* | *0.77* |
| RGB + TPI | 0.84 | 0.79 | 0.82 |
| TPI | 0.86 | 0.72 | 0.79 |
| RGB + TPI + NDVI | 0.85 | 0.72 | 0.78 |
| RGB + NDVI | 0.64 | 0.75 | 0.69 |
| RGB | 0.67 | 0.68 | 0.67 |

**Table S5.** Model testing validation results at the burrow and tile scale for all spatial resolutions. F-score = F1 score, calculated as the harmonic mean of precision and recall

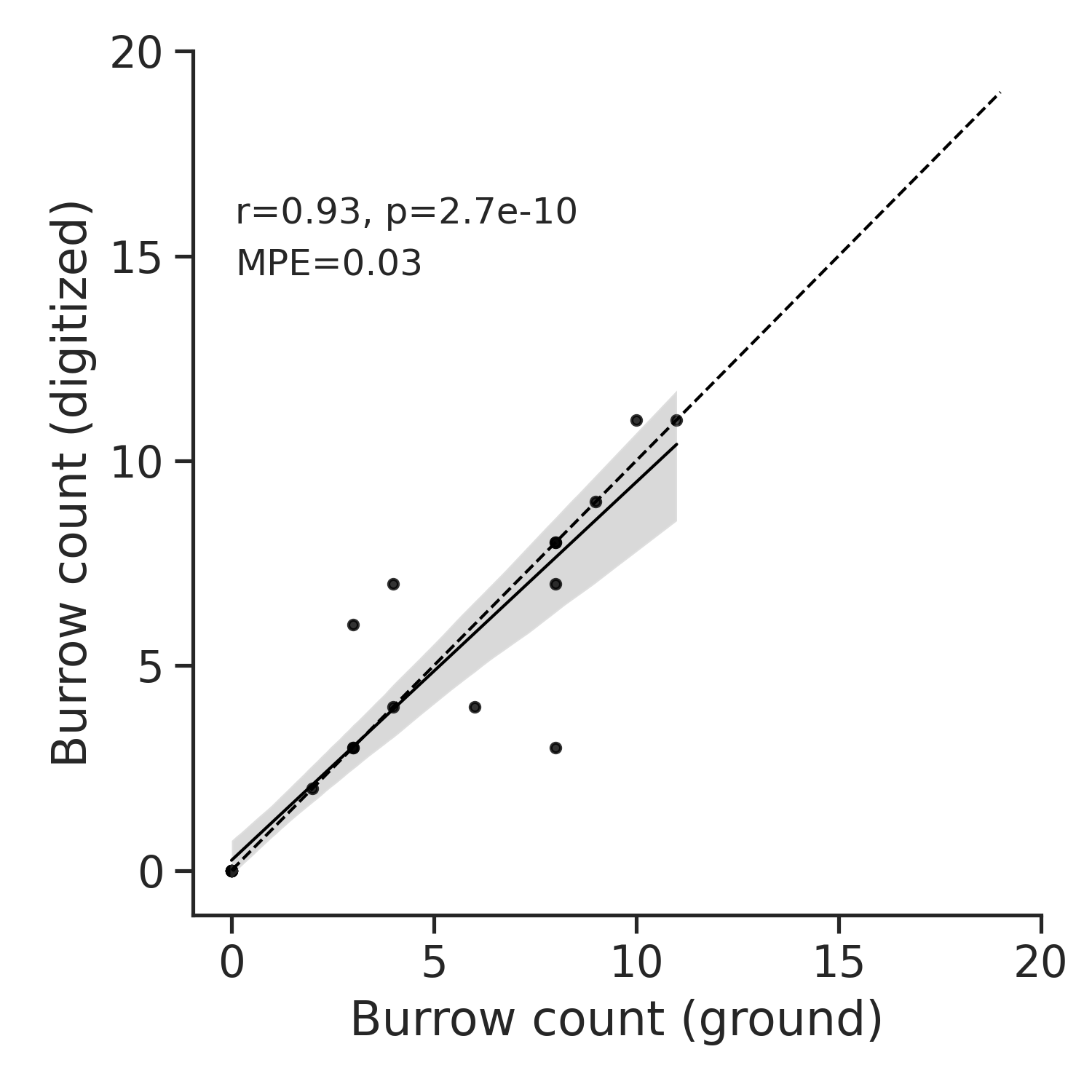
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | **Burrow scale** | | | **Tile scale** |
|  | **Inputs** | **Precision** | **Recall** | **F-score** | **Burrow count correlation** |
| **2cm** | RGB | 0.68 | 0.75 | 0.72 | 0.88 |
| **2cm** | TPI | 0.90 | 0.84 | 0.87 | 0.95 |
| **2cm** | RGB + NDVI | 0.57 | 0.76 | 0.65 | 0.83 |
| **2cm** | RGB + TPI | 0.86 | 0.84 | 0.85 | 0.95 |
| **2cm** | RGB + TPI + NDVI | 0.88 | 0.81 | 0.84 | 0.94 |
| **5cm** | RGB | 0.83 | 0.59 | 0.69 | 0.85 |
| **5cm** | TPI | 0.84 | 0.84 | 0.84 | 0.94 |
| **5cm** | RGB + NDVI | 0.85 | 0.62 | 0.72 | 0.85 |
| **5cm** | RGB + TPI | 0.87 | 0.83 | 0.85 | 0.96 |
| **5cm** | RGB + TPI + NDVI | 0.89 | 0.85 | 0.87 | 0.96 |
| **10cm** | RGB | 0.60 | 0.71 | 0.65 | 0.80 |
| **10cm** | TPI | 0.92 | 0.66 | 0.77 | 0.88 |
| **10cm** | RGB + NDVI | 0.89 | 0.42 | 0.57 | 0.73 |
| **10cm** | RGB + TPI | 0.92 | 0.69 | 0.79 | 0.93 |
| **10cm** | RGB + TPI + NDVI | 0.90 | 0.70 | 0.79 | 0.93 |
| **15cm** | RGB | 0.68 | 0.45 | 0.54 | 0.69 |
| **15cm** | TPI | 0.71 | 0.47 | 0.56 | 0.80 |
| **15cm** | RGB + NDVI | 0.69 | 0.44 | 0.54 | 0.73 |
| **15cm** | RGB + TPI | 0.92 | 0.53 | 0.67 | 0.85 |
| **15cm** | RGB + TPI + NDVI | 0.83 | 0.42 | 0.56 | 0.78 |
| **30cm** | RGB | 0.39 | 0.16 | 0.23 | 0.31 |
| **30cm** | TPI | 0.30 | 0.32 | 0.31 | 0.58 |
| **30cm** | RGB + NDVI | 0.71 | 0.05 | 0.09 | 0.26 |
| **30cm** | RGB + TPI | 0.70 | 0.21 | 0.33 | 0.51 |
| **30cm** | RGB + TPI + NDVI | 0.55 | 0.12 | 0.20 | 0.47 |

**Table S6. 5cm full colony results**

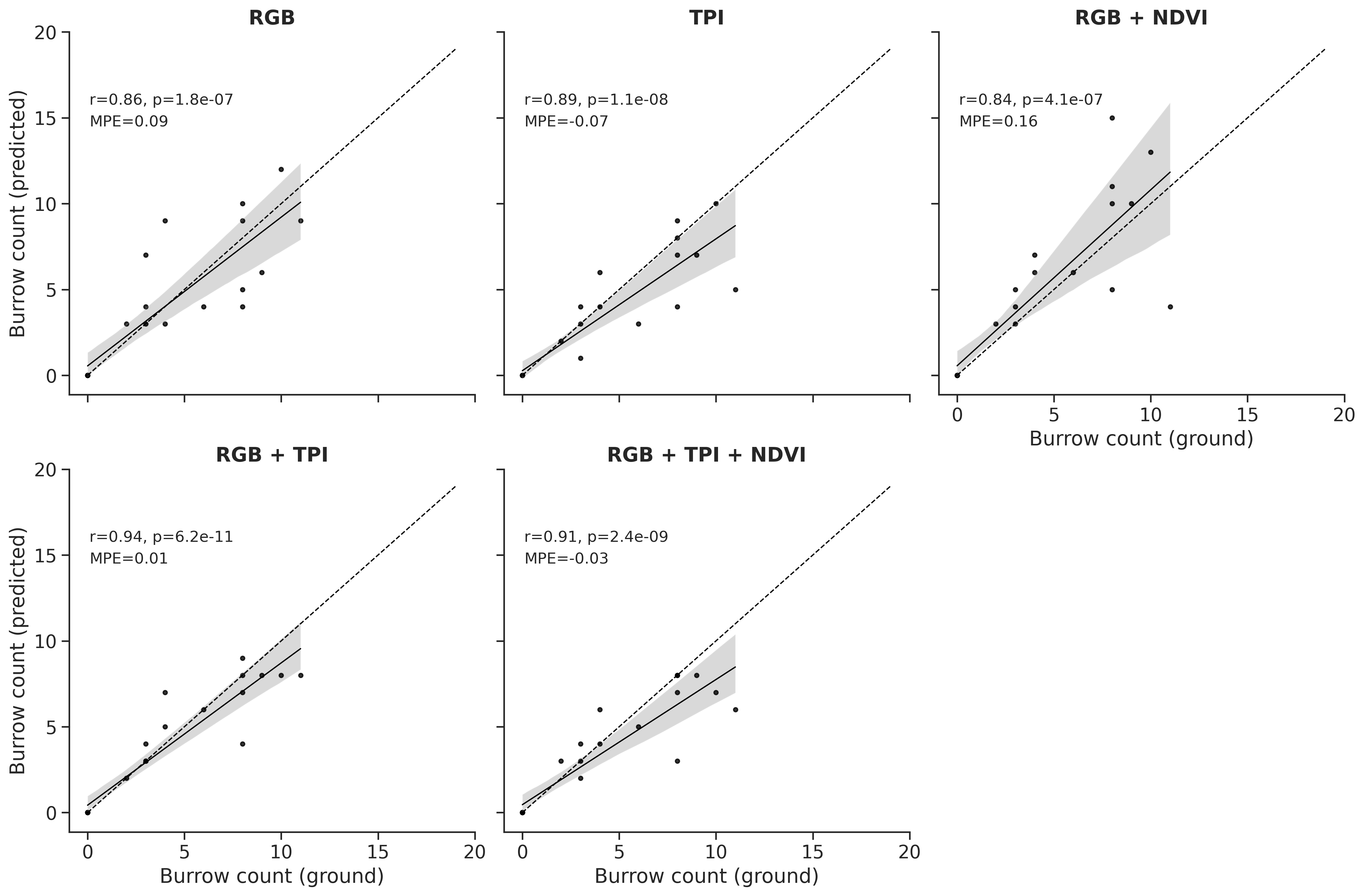
**Table S7.** Difference between model-predicted Core colony area and ground-delineated active colony area for September 2021.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Inputs** | **All pastures** | | | **Active pastures (22EW, CN, 5W)** | | |
| ***Total (ha)*** | ***Total***  ***absolute (ha)*** | ***Relative absolute*** | ***Total (ha)*** | ***Total***  ***absolute (ha)*** | ***Relative absolute*** |
| **5cm** | RGB | 8.45 | 119.11 | 0.35 | -28.11 | 82.56 | 0.24 |
| **5cm** | TPI | 169.69 | 169.69 | 0.50 | 50.35 | 50.35 | 0.15 |
| **5cm** | RGB + TPI | 250.02 | 250.02 | 0.37 | 159.76 | 159.76 | 0.24 |
| **5cm** | RGB + TPI + NDVI | 136.41 | 136.41 | 0.40 | 36.22 | 36.22 | 0.11 |
| **10cm** | RGB | 61.12 | 84.63 | 0.25 | 2.05 | 25.56 | 0.08 |
| **10cm** | TPI | -13.02 | 42.82 | 0.13 | -27.92 | 27.92 | 0.08 |
| **10cm** | RGB + TPI | -43.89 | 76.65 | 0.23 | -60.27 | 60.27 | 0.18 |
| **10cm** | RGB + TPI + NDVI | -44.28 | 77.57 | 0.23 | -60.92 | 60.92 | 0.18 |
| **15cm** | RGB | -177.52 | 208.85 | 0.62 | -193.19 | 193.19 | 0.57 |
| **15cm** | TPI | -16.47 | 62.80 | 0.19 | -27.49 | 51.78 | 0.15 |
| **15cm** | RGB + TPI | -118.30 | 135.02 | 0.40 | -126.66 | 126.66 | 0.38 |
| **15cm** | RGB + TPI + NDVI | -262.69 | 284.67 | 0.84 | -273.68 | 273.68 | 0.81 |

**Supplementary Figures**



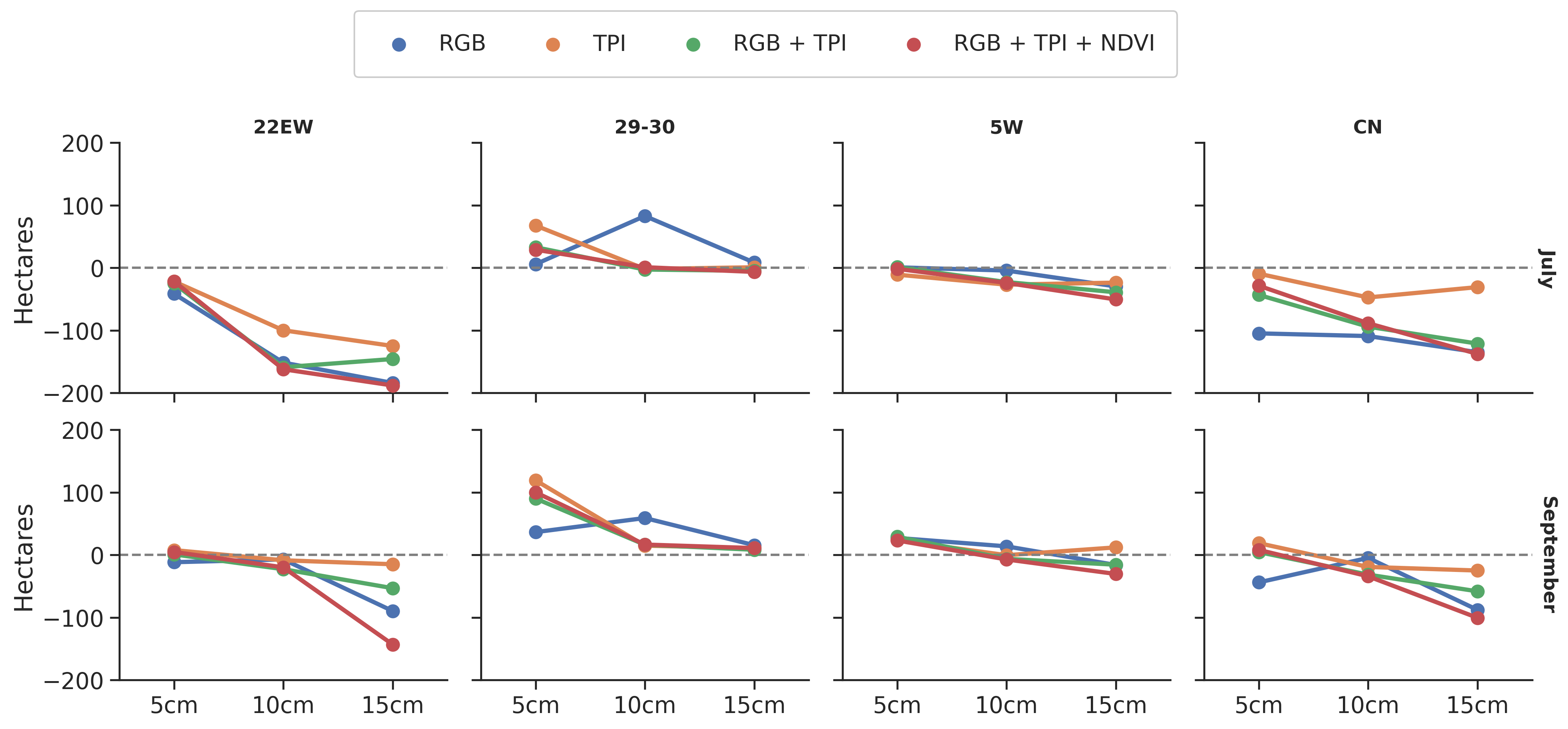
**Figure S1.** Correlation between manually digitized and ground geolocated burrow densities at the scale of 30 x 30 m testing tiles (n = 13). ‘r’ is the Pearson correlation coefficient, ‘p’ is the alpha (significance) of the correlation and ‘MPE’ is the mean percent error of predictions, an indication of bias.



**Figure S2.** Correlation between model predicted and ground geolocated burrow densities at the scale of 30 x 30 m testing tiles (n = 13). ‘r’ is the Pearson correlation coefficient, ‘p’ is the alpha (significance) of the correlation and ‘MPE’ is the mean percent error of predictions, an indication of bias.



**Figure S3.** Jaccard score (J-score) indicating overlap of model-predicted and ground delineated prairie dog colonies at increasing coarse spatial resolutions of models. Results are shown for each pasture (columns), season (rows) and combination of model inputs (colors). The horizontal line indicates a J-score of 0.70, which could be considered a high level of overlap.

**Figure S4.** Difference in model predicted vs. Ground delineated colony area (in hectares) increasing coarse spatial resolutions of models. Results are shown for each pasture (columns), season (rows) and combination of model inputs (colors). The horizontal line indicates a difference of zero hectares, which indicates equal area predicted by each approach, but does not necessarily indicate perfect overlap.